Syntactic Tagsets Affect Parsing Efficiency

Katalin Ilona Simkó[1,2], Viktória Kovács[2], Veronika Vincze[1,3]

[1]University of Szeged, Department of Informatics
Szeged, Árpád tér 2.
simko@hung.u-szeged.hu
[2]University of Szeged, Department of General Linguistics
Szeged, Egyetem u. 2.
viki921015@hotmail.com
[3]MTA-SZTE Research Group on Artificial Intelligence
Szeged, Tisza Lajos körút 103.
vinczev@inf.u-szeged.hu

Nowadays, the choice between different syntactic frameworks is getting bigger and bigger not only in theoretical, but in computational linguistics. For Hungarian, multiple representations are available for different syntactic frameworks.

In this paper, we present our findings on using different dependency labelsets in the syntactic analysis. We investigated the effect of different labelsets on the syntactic analysis itself and in applications using the syntactic parsing.

The study is based on Universal Dependencies for Hungarian [1] and in our investigation, we look at labels of adverbials, subordinating clauses and function words.

For evaluation of the syntactic parsing, we use labeled and unlabeled attachment scores as well as F-scores achieved on the content words. We believe that the syntactic parse itself is only a preprocessing step for other applications, so we try our representations (labelsets) in an NLP task classifying texts for mild cognitive impairment [2].

Our representations show improvement for syntactic parsing as well as significant improvement in the mild cognitive impairment task.

Our paper shows that choosing an appropriate syntactic representation has a powerful effect on the results of any syntax-dependent NLP application.

References

1. Vincze, V., Farkas, R., Simkó, K.I., Szántó, Zs., Varga, V.: Univerzális dependencia és morfológia magyar nyelvre. In: XII. Magyar Számítógépes Nyelvészeti Konferencia, Szeged (2015) 322–329
2. Vincze, V., Gosztolya, G., Tóth, L., Hoffmann, I., Szatlóczki, G., Bánréti, Z., Pákáski, M., Kálmán, J.: Detecting mild cognitive impairment by exploiting linguistic information from transcripts. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Berlin, Germany, Association for Computational Linguistics (2016) 181–187